ON THE PERFORMANCE OF OBJECTIVE VIDEO QUALITY METRICS FOR UHD VIDEOS

Jan Kufa

Doctoral Degree Programme (4), FEEC BUT

E-mail: xkufaj00@stud.feec.vutbr.cz

Supervised by: Ladislav Polák

E-mail: polakl@feec.vutbr.cz

Abstract: This paper deals with the performance evaluation of different objective metrics, used for the assessment of Ultra HD (UHD) video quality. Video sequences are compressed by HEVC video encoder. Objective scores are compared to each other and verified by scores from subjective tests. Such a comparison is completed by a simple correlation analysis. Differences between the objective scores obtained from the metrics implemented in the HEVC encoder and independent tool are explored in the second part of this paper. The PSNR metric is considered in this case.

Keywords: HEVC, UHD, Objective metric, Subjective test, MOS

1 INTRODUCTION

From the viewpoint of the assessment of the video quality in multimedia services [1], using of different objective metrics and subjective methods is essential [2], [3]. Objective metrics, based on mathematical approaches, are able to obtain information about the video quality in a relatively short time. Scores from subjective methods in comparison with scores from objective metrics are more adequate. However, they are expensive and not effective from the time consumption point of view. Therefore there is a big effort to develop advanced objective metrics with scores having a high correlation with subjective scores. The main aim of this paper is to compare performances of three established objective metrics with three advanced objective metrics for the evaluation of the UHD video quality, encoded by High Efficiency Video Coding (HEVC) compression technology. Comparison is completed by the subjective scores and with corresponding correlation analysis.

2 OBJECTIVE METRICS

Six objective metrics (3 established and 3 advanced) are considered altogether, in this article. The PSNR and SSIM metrics were calculated by the VQMT software. The metrics ST-MAD, VSNR and NQM were calculated in program MeTrix MuX [5]. The BVQM software was used to calculate the VQM metric [4].

- ***Peak Signal-to-Noise Ratio (PSNR)***
  It is one of the simplest and most widely used pixel-oriented metric for image and video quality evaluation purposes [2]. The higher is the PSNR value, the higher is the image quality.
- ***Structural Similarity (SSIM)***
  It is a full reference objective metric, based on the Human Visual System (HVS), which measures the similarity between two images [2]. The function is based on the fact that human eyesight is more sensitive to relative changes in brightness than to absolute changes. Value of this metrics is between 0 (low similarity) and 1 (high similarity).
- ***Video Quality Metric (VQM)***
  The VQM metric is fully based on the HVS system and compares the compressed video with the uncompressed video [2]. Index can be between 1 (low quality) and 0 (high quality).

- ***Spatiotemporal Most Apparent Distortion (ST-MAD)***
  This advanced algorithm using Most Apparent Distortion (MAD) method to estimate spatial and motion-based distortions in the video. Its lower value indicates a higher quality of the compressed video sequence [4].

- ***Visual Signal-to-Noise Ratio (VSNR)***
  The VSNR video metric quantifies the visual fidelity of distorted images. It calculates the contrast thresholds, defined as a disturbing against the reference picture with using HVS masking. If the disturbing is evaluated as an over-threshold, then the analysis continues with the perception of the low-level contrast [4]. Lower VSNR value means lower video quality.

- ***Noise Quality Measurement (NQM)***
  The NQM metric is based on the phenomenon that the psycho-visual effects of filtering and noise are separate. Its value is calculated from the measure of the frequency distortion and additive [4] noise. The higher is the NQM value, the higher is the video quality.

3 REALIZATION OF THE SUBJECTIVE TESTS

3.1 VIDEO SEQUENCES

Three short uncompressed raw video sequences were used in this work, downloaded from [6]. More information, including spatial and temporal indexes (SI and TI), can be found in the Table 1. Video sequences were compressed by HEVC video encoding algorithm and the bitrates were 1, 2, 4, 8 and 12 Mbps.

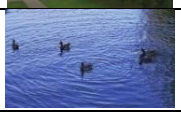
Name	Description	FPS	Resolution	Frames	SI [-]	TI [-]	Time	Thumbnail
Bospor	Floating boat on the river	30	3840x2160	300	15.9	4.4	10s	
Tree	Trees in the park	50	3840x2160	500	35.9	11.3	10s	
Duck	Ducks on the lake	50	3840x2160	500	73.6	15.7	10s	

Table 1: Parameters of the used video sequences.

3.2 SUBJECTIVE TEST SETUP

Results from the objective metrics were complemented with the scores from subjective tests. In such tests, overall 21 people were participated, after the testing of their visual acuity. All these tests were realized in controlled laboratory conditions. The presentation of video sequences was done on a computer connected to the TV. The display device was 49" LG 49UF8527. The viewing distance for all participants equaled to 104 cm (screen height 65 cm multiplied by 1.6). The Absolute Category Rating (ACR) was adopted for the subjective tests due to the lowest time duration. ACR time pattern is shown in Figure 1. The test sequences were randomly presented one at a time. At the end of every single sequence, the participant rated the quality of the video using the simple 5-point continuous scale. The range was from 0 (Bad) to 5 (Excellent). More details are described in [4].

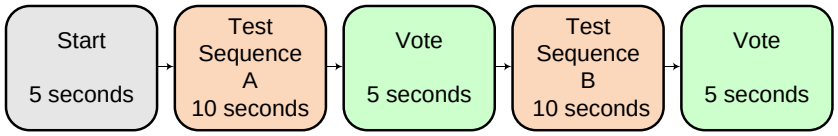


Figure 1: Time pattern for subjective tests using ACR method.

4 EXPERIMENTAL RESULTS

4.1 OBJECTIVE METRICS

The quality of the HEVC encoded UHD videos, evaluated by the above considered objective metrics, are plotted in Figure 2 to Figure 7. According to the theoretical assumptions, the higher is the bitrate, the higher is the objectives score. It is also visible that the performance of objective metrics depending on the features of the video (see Table 1). The performance of VSNR and NQM metrics is similar. Such a similarity is visible between PSNR vs. SSIM and VQM vs. ST-MAD.

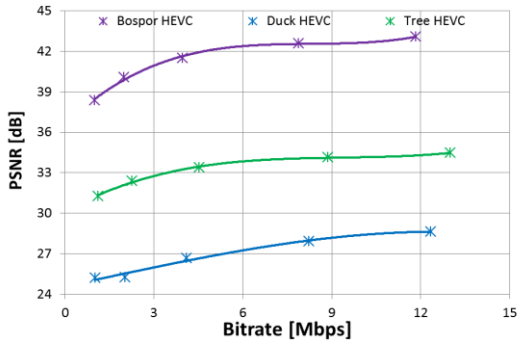


Figure 2: Dependence of PSNR on the bitrate.

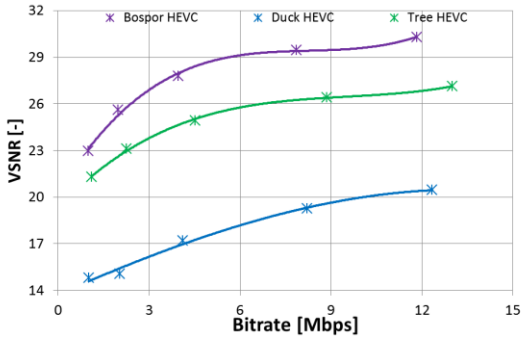


Figure 3: Dependence of VSNR on the bitrate.

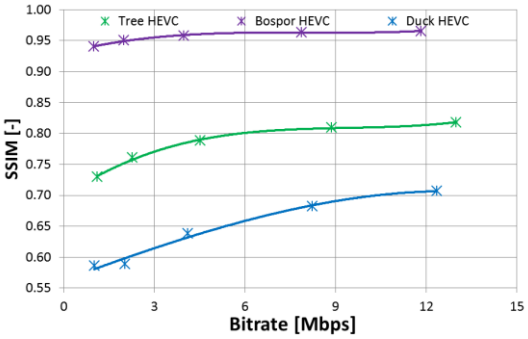


Figure 4: Dependence of SSIM on the bitrate.

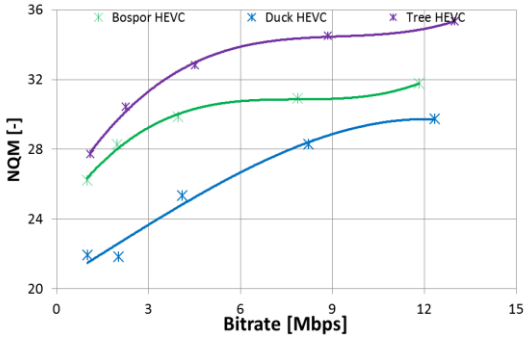


Figure 5: Dependence of NQM on the bitrate.

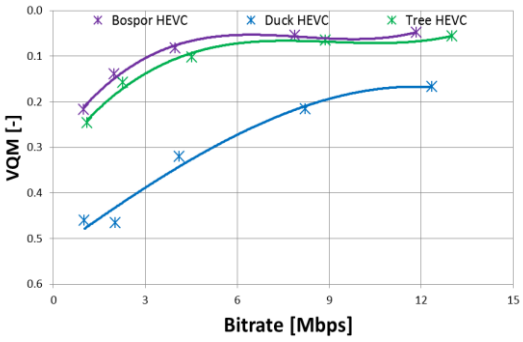


Figure 6: Dependence of VQM on the bitrate.

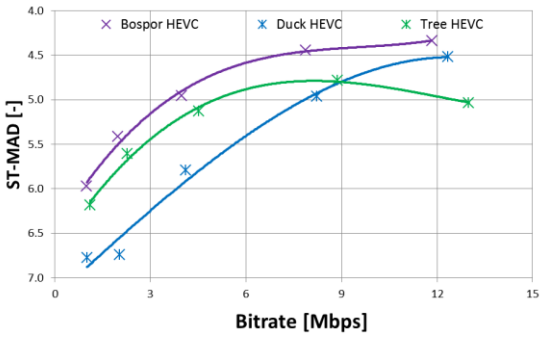


Figure 7: Dependence of ST-MAD on the bitrate.

4.2 SUBJECTIVE TEST

The subjective results from ACR method were processed and the Mean Opinion Scores (MOS) together with 95 % Confidence Interval (CI) were obtained (see Figure 8). In general, the MOS scores correspond to objectively measured quality. For the video "Duck", the obtained MOS curve is almost linear. It is an interesting fact that video "Tree" from bitrate 4 Mbps has slightly higher

MOS scores than video “Bospor”, which is probably caused by the properties of the video sequence.

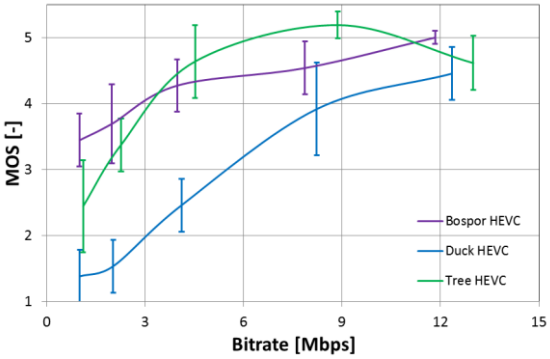


Figure 8: MOS values and 95 % and CI intervals obtained by ACR method for UHD videos.

4.3 CORRELATION OF OBJECTIVE METRICS AND SUBJECTIVE TEST

To evaluate the correlation between the objective and subjective scores, the Pearson Correlation Coefficient (PCC) was computed [1]. The outputs of such a correlation analysis are clearly presented in the Table 2. The PCC scores are between +1 and -1, where -1 and +1 mean total positive and negative linear correlation respectively, and 0 denotes no linear correlation. The VQM and ST-MAD objective metrics have negative values, because their lower score indicates higher video quality. From the obtained results is visible that objective and subjective scores correlate well. More precisely, the VQM, VSNR, NQM and ST-MAD metrics have the highest correlation with the subjective scores (bold values in the Table 2). The ST-MAD metric has the highest average correlation across all videos.

Video	Objective metric					
	PSNR	**SSIM**	**VQM**	**VSNR**	**NQM**	**ST-MAD**
Bospor	0.949	0.948	-0.933	0.972	**0.973**	-0.969
Tree	0.924	0.922	-0.962	0.935	0.945	**-0.995**
Duck	0.975	0.985	-0.989	**0.996**	0.993	-0.995
Average	0.949	0.952	-0.961	0.968	0.970	**-0.986**

Table 2: Correlation between subjective and objective metrics

4.4 PSNR OBJECTIVE METRIC CALCULATION IN THE ENCODER AND INDEPENDENT PROGRAM

Nowadays available video encoders can calculate the PSNR values during the encoding of the video sequences. Compared to professional video quality measurement tools, such a calculation can be less accurate because the encoder tries to display a better value to look more efficient. Figure 9 shows a comparison of PSNR values obtained from the HEVC encoder and VQMT tool [4]. The obtained PSNR versus bitrate curves show that the difference between the PSNR values is not higher than 2 dB.

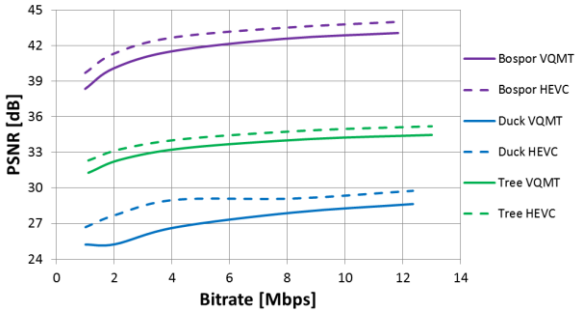


Figure 9: Comparison of PSNR metrics obtained by HEVC encoder and VQMT program.

5 CONCLUSION

Performances of different objective metrics were explored in this paper to estimate the quality of HEVC encoded UHD videos. Objectives scores were extended with scores from subjective tests. The obtained results, according to the PCC computation, show that VSNR, VQM, NQM and ST-MAD objective scores have a good correlation with the subjective scores. The ST-MAD metric has the highest average correlation across all videos. The PSNR metric has the worst correlation. This metric was developed to compare image quality. The obtained MOS for the video “Duck” was almost linearly increasing with set higher bitrate. In the case of videos “Bospor” and "Tree" were not so high increase in quality at higher bitrates. PSNR values estimated by the HEVC encoder and calculated by a professional tool were compared in next part. This comparison revealed small differences (lower than 2 dB) between the PSNR values. The difference is mostly constant for different bitrates and sequences. We can conclude that the metrics built in encoder can be used for accurate comparison of video quality, but only within an HEVC encoder. The built-in metric has to be corrected (normalized) for use to compare the coding efficiency of different encoders.

ACKNOWLEDGEMENT

Research described in this paper was financed by Czech Ministry of Education in frame of National Sustainability Program under grant LO1401. For research, infrastructure of the SIX Center was used. The research published in this submission was financially supported by the Brno University of Technology Internal Grant Agency under project no. FEKT-S-17-4426.

REFERENCES

[1] KUFA, J.; POLÁK, L.; KRATOCHVÍL, T.: HEVC/H.265 vs. VP9 for Full HD and UHD Video: Is There Any Difference in QoE? In *Proceedings 58th Symposium ELMAR*. Zadar, Croatia, 2016, pp. 51-55.

[2] KUFA, J.; KRATOCHVÍL, T.: Comparison of H265 and VP9 Coding Efficiency for Full HDTV and Ultra HDTV Applications. In *Proceedings of 25th International Conference Radioelektronika 2015*. Pardubice, Czech Republic, 2015, pp. 168-171.

[3] SHEIKH, H. R; BOVIK, A. C.: Image information and visual quality. *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430-444, Feb. 2006.

[4] BRŠEL, B.: Porovnání objektivních a subjektivních metrik kvality videa pro Ultra HDTV videosekvence. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2016. 51s. Diplomová práce. Vedoucí práce Ing. Jan Kufa

[5] GAUBATZ, M., MeTrix MuX 1.1 [online]. *Cornell University*. cit. [2018-02-03]. Available from: http://foulard.ece.cornell.edu/gaubatz/metrix_mux/#modernization

[6] Xiph.org:Derf\’s Test Media Collection. Xiph.org Video Test Media. [online]. Available from URL: https://media.xiph.org/video/derf.